# Review 185: Legendre Memory Units: Continuous-Time Representation in Recurrent Neural Networks

**Paper: https://arxiv.org/abs/2102.11417v2**

https://proceedings.neurips.cc/paper\_files/paper/2019/file/952285b9b7e7a1be5aa7849f32ffff05-Paper.pdf

המאמר הראשון בסדרה שלנו מנסה לטפל בעיה הראשונה של RNNs כלומר באי יכולת של רשתות אלו לדחוס את הזיכרון (קלט בחלון ההקשר) בצורה מספיק טובה. המאמר מציע גישה מקורית ומעניינת שמקורה במערכות דינמיות (Dynamic Systems) לבניית ייצוג הזיכרון. נניח שיש לנו פונקציית קלט רציפה (u(t ואנו רוצים לבנות מערכת ש״זוכרת את הפונקציה זו״ כלומר בונה ייצוג כך שיהיה אפשר לשחזרה באופן מדויק. תזכרו שכדי לתאר קלט דיסקרטי כמו טקסט אנו צריכים רק לעשות דיסקרטיזציה או לדגום את הפונקציה הזו.

המאמר בונה מערכת דינמית המתוארת על ידי משוואה דיפרנציאלית לינארית (מערכת דינמית, משוואה 1 במאמר) כאשר (m(t הוא וקטור הזכרון ו- (u(t כאמור הקלט (כרגע חד ממדי). מתברר שעבור בחירה מסוימת של מטריצת A במשוואה של המערכת הדינמית ניתן לתאר את הקלט (בפרק זמן מסוים) על ידי שילוב של פונקציית הזיכרון (m(t ופונקציות מתמטיות הנקראות פולינומים של Legendre (משוואה 3 במאמר). כלומר ניתן לתאר את כל מה שקרה מבחינת הקלט עד זמן מסוים על ידי פונקציה (m(t - וזה בדיוק מה שרצינו, נכון?

אולם הדאטה שלנו דיסקרטי (טוקנים נגיד) אז צריך לעשות דיסקרטיזציה (דגימה) לגישה הזו. כלומר במקום פונקציות רציפות תהיה לנו סדרת הקלט u\_t וקטור הזיכרון m\_t. גם מטריצות במערכת הדינמית שלנו צריכות לעבור דיסקרטיזציה (השערוך הרגיל של הנגזרת/גרדיאנט) ואז נקבל נוסחה רקורסיבית עבור m\_t כפונקציה של u\_t ו- m\_t-1. ניתן לתאר את את הדגימות עד t=T על ידי נוסחת נסיגה הזו.

זהו זה - יש לנו רשת בסגנון RNN כאשר הזיכרון ממודל על ידי דיסקרטיזציה של מערכת דינמית, המחשבת מקדמים של פולינומי Legendre ובאופן זה עבד לא רע אי שם ב 2020.